Online Appendix A: VB for the first-stage estimations

1 The Model

The model below is model (4) in the main body of the paper. For notational simplicity, we suppress the subscripts that indicate the model is related to the i^{th} equation of the panel SAR model.

$$Y = X\Upsilon + E, \tag{1}$$

where $vec(E) \sim N(0, \Sigma_E \otimes I_T)$. The dimensions of Y and the standardized regressors X are $T \times n$ and $T \times p$, respectively, and Υ is of dimension $p \times n$.

Let $S = E'E$ and $\Omega = \Sigma_E^{-1}$, the likelihood in equation (1) can be expressed as

$$L(\Omega) \propto |\Omega|^{\frac{T}{2}} exp\{-\frac{1}{2}tr(S\Omega)\}. \tag{2}$$

2 The Priors

2.1 The Priors of Υ

Let $\gamma = vec(\Upsilon)$. Following Bhattacharya, Pati, Pillai, and Dunson (2015), we set hierarchical DL prior for the j^{th} $(j = 1, ..., np)$ element of γ as follows:

$$\gamma_j|\phi, \tau \sim DE(\phi_j\tau), \quad \phi_j \sim Dir(a, ..., a) \tag{3}$$

Conditional prior for γ_j is $DE(\phi_j\tau)$ implies a zero mean Double Exponential or Lapalace distribution with the density $f(\gamma_j) = (2\phi_j\tau)^{-1}exp(-\frac{|\gamma_j|}{\phi_j\tau})$ for $\gamma_j \in \mathbb{R}$.

Next, we set a Gamma prior for τ:

$$\tau \sim G(npa, 1/2). \tag{4}$$

The above hierarchical DL prior for γ_j can be expressed as:

$$\gamma_j \sim N(0, \psi_j\phi_j^2\tau^2), \quad \psi_j \sim Exp(1/2) \tag{5}$$

Hence, the prior of γ is $N(\mathbf{0}, \underline{V})$ where $\underline{V} = diag(\psi_1\phi_1^2\tau^2, ..., \psi_{np}\phi_{np}^2\tau^2)$

2.2 The Priors of Ω

We set Exponential priors and DL priors for the elements of Ω

$$\begin{aligned}
&\omega_{ii} \sim Exp(\underline{s}), \quad i = 1, ..., n \\
&\omega_{ij} \sim N(0, \psi_{\omega,ij}\phi^2_{\omega,ij}\tau^2_{\omega}), \quad \psi_{\omega,ij} \sim Exp(1/2), \quad i < j = 2, ..., n, \\
&\phi_{\omega,ij} \sim Dir(a_{\omega}, ..., a_{\omega}), \quad \tau_{\omega} \sim G(\frac{n^2 - n}{2}a_{\omega}, 1/2)
\end{aligned} \tag{6}$$

where with a slight abuse of notations, we use ω_{ii} and ω_{ij} to denote the diagonal and off-diagonal elements of Ω.

3 The Posteriors

3.1 The Posteriors of Υ and its related hyper parameters

The conditional posterior of γ is $N(\tilde{\gamma}, \tilde{V})$, where

$$\begin{aligned}
&\tilde{V} = (\underline{V}^{-1} + \Omega \otimes (X'X))^{-1}, \\
&\tilde{\gamma} = \tilde{V}(\Omega \otimes X')vec(Y).
\end{aligned} \tag{7}$$

The conditional posterior of τ is $giG(npa - np, 1, \sum_{j=1}^{np} \frac{2|\gamma_j|}{\phi_j})$. [1]

The conditional posterior of ϕ_j can be derived as following: First, we have $\xi_j \sim giG(a - 1, 1, 2|\gamma_j|)$. Next let $\Xi = \sum_{j=1}^{np} \xi_j$. The conditional posterior of ϕ_j can then be found to be ξ_j/Ξ.

The conditional posterior of $1/\psi_j$ is Inverse Gaussian with mean $\sqrt{\frac{\phi_j^2 \tau^2}{\gamma_j^2}}$ and scale parameter 1.

3.2 The Posteriors of Ω and its related hyper parameters

Following Wang (2012), we focus on the last column and row of Ω then use Block Gibbs sampler to update the relevant parameters and hyper parameters.

Let H be the $n \times n$ matrix with 0 diagonal elements and the off diagonal element at i^{th} row and j^{th} column be $\psi_{\omega,ij}\phi^2_{\omega,ij}\tau^2_{\omega}$. Partition Ω, S and H as follows:

$$\Omega = \begin{pmatrix} \Omega_{-n,-n} & \omega_{-n,n} \\ \omega'_{-n,n} & \omega_{nn} \end{pmatrix} \quad S = \begin{pmatrix} S_{-n,-n} & s_{-n,n} \\ s'_{-n,n} & s_{nn} \end{pmatrix} \quad H = \begin{pmatrix} H_{-n,-n} & h_{-n,n} \\ h'_{-n,n} & h_{nn} \end{pmatrix} \tag{8}$$

where $-n$ denotes the set of all indices except for n.

[1] $y \sim giG(p, a, b)$ if $f(y) \propto y^{p-1}exp[-\frac{1}{2}(ay + b/y)]$.

Let $b_1 = \omega_{n,n} - \omega'_{-n,n}\Omega^{-1}_{-n,-n}\omega_{-n,n}$ and $b_2 = \omega_{-n,n}$. The conditional posteriors are as follows:

$$\begin{aligned}
& b_1 \sim G(\frac{T}{2}+1, \frac{s_{nn}+\underline{s}}{2}) \\
& b_2 \sim N(-C\mathbf{s}_{-n,n}, C), \quad C = ((\underline{s}+s_{nn})\Omega^{-1}_{-n,-n} + H^{*-1})^{-1}, \quad H^* = diag(h_{-n,n}) \\
& 1/\psi_{\omega,ij} \sim iG(\sqrt{\frac{\tau^2_\omega \phi^2_{\omega,ij}}{\omega^2_{ij}}}, 1) \\
& \tau_\omega \sim giG(\frac{n^2-n}{2}(a_3-1), 1, \sum_{i<j}\frac{2|\omega_{ij}|}{\phi_{\omega,ij}}) \\
& \xi_{\omega,ij} \sim giG(a_\omega - 1, 1, 2|\omega_{ij}|), \quad \phi_{\omega,ij} = \xi_{\omega,ij}/\sum_{i<j}\xi_{\omega,ij}
\end{aligned} \tag{9}$$

4 Optimal VB q densities

4.1 $q(\gamma)$

$$q(\gamma) \sim N(\bar{\gamma}, \overline{V}), \tag{10}$$

where

$$\overline{V} = (V^{-1} + \overline{\Omega} \otimes (X'X))^{-1},$$
$$\overline{\gamma} = \overline{V}(\overline{\Omega} \otimes X')vec(Y),$$

and

$$V^{-1} = diag(\frac{1}{\overline{\psi_1}\ \overline{\phi_1^2}\ \overline{\tau^2}}, ..., \frac{1}{\overline{\psi_{np}}\ \overline{\phi_{np}^2}\ \overline{\tau^2}})$$

Note that using Matlab usually it is faster to calculate $\overline{V}$ by $(V^{-1} + \overline{\Omega} \otimes (X'X))\ I_{np}$ than by directly taking the inverse of $(V^{-1} + \overline{\Omega} \otimes (X'X))$.

4.2 $q(\tau)$

$$q(\tau) \sim giG[npa - np, 1, \sum_{j=1}^{np} 2(\overline{\gamma}^2_j + \overline{V}_{jj})^{1/2}\frac{1}{\overline{\phi_j}}], \tag{11}$$

Let $\chi = \sum_{j=1}^{np} 2(\overline{b}^2_j + \overline{V}_{jj})^{1/2}\frac{1}{\overline{\phi_j}}$, we have

$$\bar{\tau} = \frac{\sqrt{\chi}K_{npa-np+1}(\sqrt{\chi})}{K_{npa-np}(\sqrt{\chi})}$$

$$\overline{\tau^2} = \bar{\tau}^2 + \chi[\frac{K_{npa-np+2}(\sqrt{\chi})}{K_{npa-np}(\sqrt{\chi})} - (\frac{K_{npa-np+1}(\sqrt{\chi})}{K_{npa-np}(\sqrt{\chi})})^2]$$

where $K_*[\bullet]$ is the modified Bessel functions of the second kind.

4.3 $q(\psi_j)$

$$q(\psi_j^{-1}) \sim iG(\sqrt{\frac{\overline{\phi_j^2}\,\overline{\tau^2}}{\overline{\gamma}_j^2 + \overline{V}_{jj}}}, 1), \tag{12}$$

Let $\rho = \sqrt{\frac{\overline{\phi_j^2}\,\overline{\tau^2}}{\overline{\gamma}_j^2+\overline{V}_{jj}}}$,

$$\overline{\psi_j^{-1}} = \rho$$

and

$$\overline{\psi}_j = 1 + 1/\rho$$

4.4 $q(\phi_j)$

$$q(\xi_j) \sim giG(a-1, 1, 2\sqrt{\overline{\gamma}_j^2 + \overline{V}_{jj}}) \tag{13}$$

Let $\varpi = 2\sqrt{\overline{\gamma}_j^2 + \overline{V}_{jj}}$, we have

$$\overline{\xi_j} = \frac{\sqrt{\varpi}K_a(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})},$$

and

$$var(\xi_j) = \varpi\{\frac{K_{a+1}(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})} - [\frac{K_a(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})}]^2\},$$

where $var(\bullet)$ denotes the variance.

Scaling ξ_i, we have

$$\overline{\phi_j} = \frac{\overline{\xi_j}}{\sum_{j=1}^{np}\overline{\xi_j}},$$

and

$$\overline{\phi_j^2} = \overline{\phi_j}^2 + \frac{var(\xi_j)}{(\sum_{j=1}^{np}\overline{\xi_j})^2}$$

Thus, the optimal q density of $\phi_{i,j}$ takes the following form:

$$q(\phi_j) \sim giG(a-1, \sum_{j=1}^{np}\overline{\xi_j}, \frac{2\sqrt{\overline{\gamma}_j^2 + (\overline{V}_{jj})^2}}{\sum_{j=1}^{np}\overline{\xi_j}}) \tag{14}$$

4.5 $q(b_1)$

$$q(b_1) \sim G(\frac{T}{2}, \overline{s}_{n,n}), \tag{15}$$

where

$$\overline{s}_{n,n} = \frac{1}{2}(s_{n,n} + tr(X'X\overline{V_n}) + \underline{s}),$$

and

$$\overline{V_n} = V_{(n-1)\times p+1:n\times p,(n-1)\times p+1:n\times p}.$$

Hence

$$\overline{b_1} = \frac{\frac{T}{2}}{\overline{s}_{n,n}}$$

4.6 $q(b_2)$

Let $\overline{s}_{-n,n} = s_{-n,n} + \tilde{s}_{-n,n}$, where $\tilde{s}_{-n,n}$ is a $(n-1)\times 1$ vector with the j^{th} element being $tr(X'XA_j)$ and $A_j = V_{(j-1)\times p+1:j\times p,(j-1)\times p+1:j\times p}$. [2]

$$q(b_2) \sim N(-\overline{C}\overline{s}_{-n,n}, \overline{C}), \tag{16}$$

where

$$\overline{C} = (2\overline{s}_{n,n}\Omega^{-1}_{-n.-n} + \overline{H}^{*-1})^{-1},$$

and

$$\overline{b_2} = (-\overline{C}\overline{s}_{-n,n}).$$

Note that $\overline{H}^* = diag(\overline{h}_{-n,n})$, and j^{th} element of $\overline{h}_{-n,n}$ is $\overline{\psi_{\omega,jn}}\overline{\phi^2_{\omega,jn}}\overline{\tau^2_\omega}$.

4.7 $q(\tau_\omega)$

$$q(\tau_\omega) \sim giG[\frac{n^2-n}{2}(a_\omega - 1), 1, \sum_{j<k} 2(\overline{\omega_{jk}}^2 + \overline{C}_{jj})^{1/2}\frac{1}{\overline{\overline{\phi}_{\omega,jk}}}],^{3} \tag{17}$$

Let $\chi_\omega = \sum_{j<k} 2(\overline{\omega_{jk}}^2 + \overline{V}_{jj})^{1/2}(\overline{\phi_{\omega,jk}})^{-1}$, we have

$$\overline{\tau_\omega} = \frac{\sqrt{\chi_\omega}K_{\frac{n^2-n}{2}(a_\omega-1)+1}(\sqrt{\chi_\omega})}{K_{\frac{n^2-n}{2}(a_\omega-1)}(\sqrt{\chi_\omega})}$$

$$\overline{\tau^2_\omega} = \overline{\tau_\omega}^2 + \chi_\omega[\frac{K_{\frac{n^2-n}{2}(a_\omega-1)+2}(\sqrt{\chi})}{K_{\frac{n^2-n}{2}(a_\omega-1)}(\sqrt{\chi_\omega})} - (\frac{K_{\frac{n^2-n}{2}(a_\omega-1)+1}(\sqrt{\chi_\omega})}{K_{\frac{n^2-n}{2}(a_\omega-1)}(\sqrt{\chi_\omega})})^2]$$

[2] To calculate $\overline{s}_{-i,i}$, for $i \neq n$, we need to delete the relevant $(i-1)\times p+1$ to $i \times p$ rows and $(i-1)\times p+1$ to $i\times p$ columns of $\overline{V}$ to construct A_j.

[3] Note that $\overline{C}_{jj}$ changes when k changes.

4.8 $q(\psi_{\omega,jn})$

$$q(\psi_{\omega,jn}^{-1}) \sim iG(\sqrt{\frac{\overline{\phi_{\omega,jn}^2}\,\overline{\tau_\omega^2}}{\overline{\omega_{jn}^2}+\overline{C}_{jj}}}, 1), \tag{18}$$

Let $\rho_\omega = \sqrt{\frac{\overline{\phi_{\omega,jn}^2}\,\overline{\tau_\omega^2}}{\overline{\omega_{jn}}^2+\overline{C}_{jj}}}$,

$$\overline{\psi_{\omega,jn}^{-1}} = \rho_\omega$$

and

$$\overline{\psi_{w,jn}} = 1 + 1/\rho_\omega$$

4.9 $q(\phi_{\omega,jn})$

$$q(\xi_{\omega,jn}) \sim giG(a_\omega - 1, 1, 2\sqrt{\overline{\omega_{jn}}^2 + \overline{C}_{jj}}) \tag{19}$$

Let $\varpi_\omega = 2\sqrt{\overline{\omega_{jn}}^2 + \overline{C}_{jj}}$, we have

$$\overline{\xi_{\omega,jn}} = \frac{\sqrt{\varpi_\omega}K_{a_\omega}(\sqrt{\varpi_\omega})}{K_{a_\omega-1}(\sqrt{\varpi_\omega})},$$

and

$$var(\xi_{\omega,jn}) = \varpi_\omega\{\frac{K_{a_\omega+1}(\sqrt{\varpi_\omega})}{K_{a_\omega-1}(\sqrt{\varpi_\omega})} - [\frac{K_{a_\omega}(\sqrt{\varpi_\omega})}{K_{a_\omega-1}(\sqrt{\varpi_\omega})}]^2\},$$

where $var(\bullet)$ denotes the variance.

Scaling $\xi_{\omega,jn}$, we have

$$\overline{\phi_{\omega,jn}} = \frac{\overline{\xi_{\omega,jn}}}{\sum_{j<k}\overline{\xi_{\omega,jk}}},$$

and

$$\overline{\phi_{\omega,jn}^2} = \overline{\phi_{\omega,jn}}^2 + \frac{var(\xi_{\omega,jn})}{(\sum_{j<k}\overline{\xi_{\omega,jk}})^2}$$

Thus, the optimal q density of $\phi_{\omega,ij}$ takes the following form:

$$q(\phi_{\omega,jn}) \sim giG(a-1, \sum_{j<k}\overline{\xi_{\omega,jk}}, \frac{2\sqrt{\overline{\omega_{jn}}^2 + (\overline{C}_{jj})^2}}{\sum_{j<k}\overline{\xi_{\omega,jk}}}) \tag{20}$$

5 ELBO

The evidence lower bound ($ELBO$) is as follows:[4]

$$\begin{aligned}ELBO &= E\{\log p(Y,\gamma,\tau,\psi,\phi,b_1,b_2,\tau_\omega,\psi_\omega,\phi_\omega)\} - E\{\log q(\gamma,\tau,\psi,\phi,b_1,b_2\tau_\omega,\psi_\omega,\phi_\omega)\}\\ &= E\{\log p(Y|\gamma,\tau,\psi,\phi,b_1,b_2\tau_\omega,\psi_\omega,\phi_\omega)\} + E\{\log p(\gamma)\} + \sum E\{\log p(b_1)\} + \sum E\{\log p(b_2)\}\\ &+ E\{\log p(\phi)\} + E\{\log p(\psi)\} + E\{\log p(\tau)\} + E\{\log p(\phi_\omega)\} + E\{\log p(\psi_\omega)\} + E\{\log p(\tau_\omega)\}\\ &- E\{\log q(\gamma)\} - \sum E\{\log q(b_1)\} - \sum E\{\log q(b_2)\} - E\{\log q(\phi)\} - E\{\log q(\psi)\} - E\{\log q(\tau)\}\\ &- E\{\log q(\phi_\omega)\} - E\{\log q(\psi_\omega)\} - E\{\log q(\tau_\omega)\}\end{aligned} \tag{21}$$

where

$$\begin{aligned}&E\{\log p(Y|\gamma,\tau,\psi,\phi,b_1,b_2,\tau_\omega,\psi_\omega,\phi_\omega)\} =\\ &-\frac{T}{2}log|\overline{\Sigma}| - \frac{1}{2}\left[vec(\overline{\Omega})\right]^{'} vec(Y^{'}Y - 2XBY^{'} + XBB^{'}X^{'} + \sum_{j=1}^{n} XK_jX^{'}) + const,\end{aligned} \tag{22}$$

where the elements of K_j are retrieved from appropriate rows and columns of $\overline{V}$,

$$E\{\log p(\gamma)\} = -\frac{1}{2}E[log(|V|)] - \frac{1}{2}\left[\overline{\gamma}'V^{-1}\overline{\gamma} + tr(V^{-1}\overline{V})\right] + const, \tag{23}$$

where $E[log(|V|)] = \sum_{j=1}^{np} E[\log(\psi_j)+2\log(\phi_j)]+2npE[\log(\tau)]$, $E[\log(\psi_j)+2\log(\phi_j)] = \int_0^\infty q(\psi_j)\log(\psi_j)d\psi_j + 2\int_0^\infty q(\phi_j)\log(\phi_j)d\phi_j$ and $E[\log(\tau)] = \int_0^\infty q(\tau)\log(\tau)d\tau$.

$$E\{\log p(b_1)\} = -(\underline{\nu}-1)log(\overline{s}_{n,n}) - \underline{s}\overline{b_1} + const, \tag{24}$$

$$E\{\log p(b_2)\} = -\frac{1}{2}E(log(|H^*|)) - \frac{1}{2}\left[\overline{b_2}'H^*\overline{b_2} + tr('H^{*-1}\overline{C})\right] + const, \tag{25}$$

where $E[log(|H^*|)] = \sum_{j=1}^{n-1} E[\log(\psi_{\omega,ij})+2\log(\phi_{\omega,ij})]+2(n-1)E[\log(\tau_\omega)]$, $E[\log(\psi_{\omega,ij})+2\log(\phi_{\omega,ij})] = \int_0^\infty q(\psi_{\omega,ij})\log(\psi_{\omega,ij})d\psi_{\omega,ij} + 2\int_0^\infty q(\phi_{\omega,ij})\log(\phi_{\omega,ij})d\phi_{\omega,ij}$ and $E[\log(\tau_\omega)] = \int_0^\infty q(\tau_\omega)\log(\tau_\omega)d\tau_\omega$.

$$E\{\log p(\tau)\} = -npa\left[\int_0^\infty (q(\tau)\log\tau)\, d\tau\right] - 0.5\overline{\tau} + const, \tag{26}$$

$$E\{\log p(\psi)\} = -0.5\sum_{j=1}^{np}\overline{\psi_j} + const, \tag{27}$$

$$E\{\log p(\phi)\} = -(a-1)\sum_{j=1}^{np}\left[\int_0^\infty (q(\phi_j)\log\phi_j)d\phi_j\right] \tag{28}$$

[4] It is seen that the $ELBO$ described below contains many constant terms. Those terms need to be dropped from $ELBO$s in VB iterations to speed up the process.

$$E\{\log p(\tau_\omega)\} = -\frac{n^2-n}{2}a_\omega \left[\int_0^\infty (q(\tau_\omega)\log \tau_\omega)\, d\tau_\omega\right] - 0.5\left[\int_0^\infty (q(\tau_\omega)\tau_\omega)\, d\tau_\omega\right] + const, \tag{29}$$

$$E\{\log p(\psi_\omega)\} = -0.5 \sum_{j=1}^{(n^2-n)/2} \int_0^\infty (\psi_{\omega,j} q(\psi_{\omega,j})) d\psi_{\omega,j} + const, \tag{30}$$

$$E\{\log p(\phi_\omega)\} = -(a_\omega - 1) \sum_{j=1}^{(n^2-n)/2} \left[\int_0^\infty (q(\phi_{\omega,j})\log \phi_{\omega,j}) d\phi_{\omega,j}\right] \tag{31}$$

$$E\{\log q(\gamma)\} = -\frac{1}{2} log(|\overline{V}|) + const, \tag{32}$$

$$E\{\log p(b_1)\} = log(\overline{s}_{n,n}) + const, \tag{33}$$

$$E\{\log p(b_2)\} = -\frac{1}{2} E(log(\overline{C})) + const, \tag{34}$$

$$E\{\log q(\tau)\} = \int_0^\infty q(\tau)\log q(\tau) d\tau. \tag{35}$$

$$E\{\log q(\psi)\} = \sum_{j=1}^{np} \int_0^\infty q(\psi_j)\log q(\psi_j) d\psi_j \tag{36}$$

$$E\{\log q(\phi)\} = \sum_{j=1}^{np} \int_0^\infty q(\phi_j)\log q(\phi_j) d\phi_j \tag{37}$$

$$E\{\log q(\tau_\omega)\} = \int_0^\infty q(\tau_\omega)\log q(\tau_\omega) d\tau_\omega. \tag{38}$$

$$E\{\log q(\psi_\omega)\} = \sum_{j=1}^{(n^2-2)/2} \int_0^\infty q(\psi_{\omega,j})\log q(\psi_{\omega,j}) d\psi_{\omega,j} \tag{39}$$

$$E\{\log q(\phi_\omega)\} = \sum_{j=1}^{(n^2-2)/2} \int_0^\infty q(\phi_{\omega,j})\log q(\phi_{\omega,j}) d\phi_{\omega,j} \tag{40}$$

Online Appendix B: VB for the second-stage estimations

1 The Model

For notational simplicity, we suppress the subscripts denoting the i^{th} variable in model (5) of the main paper. In addition, we use Z and θ to denote the matrix of the regressors and the vector of parameters used in that model.

$$y = Z\theta + u \tag{1}$$

where y is $T \times 1$ vector of responses, Z is the $T \times k$ matrix of standardized regressors, u is the $T \times 1$ vector of $i.i.d$ Normal errors with mean 0 and unknown variance σ^2.

2 The Priors

Following Bhattacharya, Pati, Pillai, and Dunson (2015) (BPPD), we elicit hierarchical DL prior as follows:[1]

$$\theta_j|\phi, \tau \sim DE(\phi_j \tau), \quad \phi_j \sim Dir(a, ..., a) \tag{2}$$

Conditional prior for θ_j is $DE(\phi_j \tau)$ implies a zero mean Double Exponential or Lapalace distribution with the density $f(\theta_j) = (2\phi_j\tau)^{-1} exp(-\frac{|\theta_j|}{\phi_j \tau})$ for $\theta_j \in \mathbb{R}$.

Next, we set a Gamma priors for τ and σ^{-2}:

$$\tau \sim G(ka, 1/2), \quad \sigma^{-2} \sim G(\nu, S). \tag{3}$$

[1]In the main paper, we use $\tilde{\bullet}$ to denote a hyperparameter $\bullet$ to distinguish it from that set for the first-stage estimation. For notational simplicity, we drop $\tilde{}$ on the hyperparameters in this appendix.

The above hierarchical DL prior for θ_j can be expressed as:

$$\theta_j \sim N(0, \psi_j \phi_j^2 \tau^2), \quad \psi_j \sim Exp(1/2) \tag{4}$$

Hence, the prior of θ is $N(\mathbf{0}, V)$ where $V = diag(\psi_1 \phi_1^2 \tau^2, ..., \psi_k \phi_k^2 \tau^2)$.

3 The Posteriors

Multiply the likelihood and the priors, we have

$$\begin{aligned} & L(y|\theta, \sigma^2) p(\theta|\phi, \tau, \sigma^2) p(\tau) p(\phi) p(\sigma^2) \\ \propto & \frac{1}{(\sqrt{\sigma^2})^T} exp[-\frac{1}{2\sigma^2}(y - Z\theta)'(y - Z\theta)] \times \prod_{j=1}^{k} \frac{1}{2\phi_j \tau} exp[-\sum_{j=1}^{k} \frac{|\theta_j|}{\phi_j \tau}] \times \\ & \tau^{ka-1} exp(-\frac{\tau}{2}) \times \prod_{j=1}^{k} \phi_j^{a-1} \times (\sigma^{-2})^{\nu-1} exp(-S/\sigma^2) \end{aligned} \tag{5}$$

or

$$\begin{aligned} & L(y|\theta, \sigma^2) p(\theta|\phi, \tau, \psi, \sigma^2) p(\tau) p(\phi) p(\psi) p(\sigma^2) \\ \propto & \frac{1}{(\sqrt{\sigma^2})^T} exp[-\frac{1}{2\sigma^2}(y - Z\theta)'(y - Z\theta)] \times \prod_{j=1}^{k} \frac{1}{\sqrt{\psi_j \phi_j^2 \tau^2}} exp[-\sum_{j=1}^{k} \frac{\theta_j^2}{2\psi_j \phi_j^2 \tau^2}] \times \\ & \tau^{ka-1} exp(-\frac{\tau}{2}) \times \prod_{j=1}^{k} \phi_j^{a-1} \times \prod_{j=1}^{k} exp(-\psi_j/2) \times (\sigma^{-2})^{\nu-1} exp(-S/\sigma^2) \end{aligned} \tag{6}$$

In (6), the items involving θ are $exp[-\frac{1}{2\sigma^2}(y - Z\theta)'(y - Z\theta)] \times \prod_{j=1}^{k} \frac{1}{\sqrt{\psi_j \phi_j^2 \tau^2}} exp[-\sum_{j=1}^{k} \frac{\theta_j^2}{2\psi_j \phi_j^2 \tau^2}]$. Thus, the conditional posterior of θ is $N(\bar{b}, \overline{V})$, with $\overline{V} = [\sigma^{-2} Z'Z + V^{-1}]^{-1}$ and $\bar{b} = \sigma^{-2} \overline{V} Z' y$.

In (5), the items involving τ are $\frac{1}{\tau^k} exp[-\sum_{j=1}^{k} \frac{|\theta_j|}{\phi_j \tau}] \times \tau^{ka-1} exp(-\frac{\tau}{2})$, which can be written as $\tau^{ka-k-1} exp[-\frac{1}{2}(\tau + \frac{1}{\tau}(\sum_{j=1}^{k} \frac{2|\theta_j|}{\phi_j}))]$. Thus the conditional posterior of τ is $giG(ka - k, 1, \sum_{j=1}^{k} \frac{2|\theta_j|}{\phi_j})$.[2]

To find the conditional posteriors of ϕ_j, we integrate τ out following BPPD. Collecting the terms involving ϕ_j, we have $\prod_{j=1}^{k} (\phi_j^{a-1} \frac{1}{\phi_j}) \int_{\tau=0}^{\infty} \tau^{ka-k-1} exp(-\frac{\tau}{2}) exp[-\sum_{j=1}^{k} \frac{|\theta_j|}{\phi_j \tau}] d\tau$. Thus, we can

[2] $y \sim giG(p, a, b)$ if $f(y) \propto y^{p-1} exp[-\frac{1}{2}(ay + b/y)]$.

derive the conditional posterior of ϕ_j as following: First, we have $\xi_j \sim giG(a-1, 1, 2|\theta_j|)$. Next let $\Xi = \sum_{j=1}^{k} \xi_j$. The conditional posterior of ϕ_j can then be found to be ξ_j/Ξ.

In (6), the terms involving ψ_js are $\prod_{j=1}^{k} \psi_j^{-\frac{1}{2}} exp[-\sum_{j=1}^{k} \frac{\theta_j^2}{2\psi_j \phi_j^2 \tau^2}] \times \prod_{j=1}^{k} exp(-\psi_j/2)$. Thus the conditional posterior of $1/\psi_j$ is Inverse Gaussian with mean $\sqrt{\frac{\phi_j^2 \tau^2}{\theta_j^2}}$ and scale parameter 1.

In (6), the terms involving σ^{-2} are

$$\frac{1}{(\sqrt{\sigma^2})^T} exp[-\frac{1}{2\sigma^2}(y - Z\theta)'(y - Z\theta)] \times (\sigma^{-2})^{\nu-1} exp(-S/\sigma^2)$$

Thus the conditional posterior of σ^{-2} is $G(\frac{T}{2} + \nu, \frac{1}{2}(y - Z\theta)'(y - Z\theta) + S)$.

4 Variational Bayes

The optimal VB q densities are as follows:

4.1 $q(\theta)$

$$q(\theta) \sim N(\overline{\theta}, \overline{V}), \tag{7}$$

where

$$\overline{V} = (\frac{\frac{T}{2} + \nu}{\overline{S}} Z'Z + V^{-1})^{-1}$$

$$\overline{\theta} = (\frac{\frac{T}{2} + \nu}{\overline{S}}) \overline{V} Z'y$$

$$V^{-1} = diag(\overline{\psi_1^{-1}}\ \overline{\phi_1^{-2}}\overline{\tau^{-2}}, ..., (\overline{\psi_k^{-1}}\ \overline{\phi_k^{-2}}\overline{\tau^{-2}})$$

$$E[\log(|V^{-1}|)] = \sum_{j=1}^{k} E[\log(\psi_j) + 2\log(\phi_j)] + 2kE[\log(\tau)],$$

where

$$E[\log(\psi_j) + 2\log(\phi_j)] = \int_0^\infty q(\psi_j) \log(\psi_j) d\psi_j + 2\int_0^\infty q(\phi_j) \log(\phi_j) d\phi_j$$

and

$$E[\log(\tau)] = \int_0^\infty q(\tau)\log(\tau)d\tau$$

4.2 $q(\sigma^{-2})$

$$q(\sigma^{-2}) \sim G(\frac{T}{2}+\nu, \overline{S}), \tag{8}$$

where

$$\overline{S} = \frac{1}{2}[||y - Z\overline{\theta}||^2 + tr(Z'Z\overline{V})] + S$$

Hence

$$\overline{\sigma}^{-2} = \frac{\frac{T}{2}+\nu}{\overline{S}},$$

and

$$E[\log(\sigma^{-2})] = \psi(\nu + \frac{T}{2}) - log(\overline{S})$$

where $\psi(\bullet)$ is the Digamma function.

4.3 $q(\tau)$

$$q(\tau) \sim giG[ka-k, 1, \sum_{j=1}^{k} 2(\overline{\theta}_j^2 + \overline{V}_{jj})^{1/2}\frac{1}{\overline{\phi_j}}], \tag{9}$$

Let $\chi = \sum_{j=1}^{k} 2(\overline{\theta}_j^2 + \overline{V}_{jj})^{1/2}\frac{1}{\overline{\phi_j}}$, we have

$$\overline{\tau} = \frac{\sqrt{\chi}K_{ka-k+1}(\sqrt{\chi})}{K_{ka-k}(\sqrt{\chi})}$$

and

$$\overline{\tau^2} = \overline{\tau}^2 + \chi[\frac{K_{ka-k+2}(\sqrt{\chi})}{K_{ka-k}(\sqrt{\chi})} - (\frac{K_{ka-k+1}(\sqrt{\chi})}{K_{ka-k}(\sqrt{\chi})})^2]$$

where $K_*[\bullet]$ is the modified Bessel functions of the second kind.

4.4 $q(\psi_j)$

$$q(\frac{1}{\psi_j}) \sim iG(\sqrt{\frac{\overline{\phi_j^2}\,\overline{\tau^2}}{\overline{\theta}_j^2 + \overline{V}^{jj}}}, 1), \tag{10}$$

Let $\rho = \sqrt{\frac{\overline{\phi_j^2}\,\overline{\tau^2}}{\overline{\theta}_j^2 + \overline{V}^{jj}}}$,

$$\overline{\frac{1}{\psi_j}} = \rho$$

and

$$\overline{\psi}_j = 1 + 1/\rho$$

Note that to calculate $ELBO$, we need to use the following optimal q density of ψ_j.[3]

$$q(\psi_j) = (\frac{\psi_j}{2\pi})^{1/2} \exp\{\frac{-(\psi_j - \rho)^2}{2\rho^2\psi_j}\} \tag{11}$$

4.5 $q(\phi_j)$

$$q(\xi_j) \sim giG(a-1, 1, 2\sqrt{\overline{\theta}_j^2 + (\overline{V}_{jj})^2}) \tag{12}$$

Let $\varpi = 2\sqrt{\overline{\theta}_j^2 + (\overline{V}_{jj})^2}$, we have

$$\bar{\xi}_j = \frac{\sqrt{\varpi} K_a(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})},$$

and

$$var(\xi_j) = \varpi\{\frac{K_{a+1}(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})} - [\frac{K_a(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})}]^2\}.$$

where $var(\bullet)$ denotes the variance.

Scaling ξ, we have

$$\overline{\phi}_j = \frac{\bar{\xi}_j}{\sum^k \bar{\xi}_j},$$

[3] If x is distributed as $f(x)$, then $y = 1/x$ is distributed as $\frac{1}{y^2} f(\frac{1}{y})$.

and

$$\overline{\phi_j^2} = \overline{\phi}_j^2 + \frac{var(\xi_j)}{(\sum^k \overline{\xi}_j)^2}$$

Thus, the optimal q density of ϕ_j takes the following form:

$$q(\phi_j) \sim giG[a-1, \sum^k \overline{\xi}_j, (2\sqrt{\overline{\theta}_j^2 + (\overline{V}_{jj})^2})/(\sum^k \overline{\xi}_j)] \tag{13}$$

4.6 *ELBO*

The evidence lower bound ($ELBO$) is as follows:[4]

$$\begin{aligned} ELBO &= E\{\log p(\mathbf{y}, \theta, \sigma^2, \phi, \tau, \psi)\} - E\{\log q(\theta, \sigma^2, \phi, \tau, \psi)\} \\ &= E\{\log p(\mathbf{y}|\theta, \sigma^2, \phi, \tau, \psi)\} + E\{\log p(\theta)\} + E\{\log p(\sigma^2)\} \\ &+ E\{\log p(\phi)) + E(\log p(\psi)\} + E\{\log p(\tau)\} \\ &- E\{\log(q(\theta))\} - E\{\log q(\sigma^2)\} - E\{\log q(\tau)\} - E\{\log q(\phi)\} - E\{\log q(\psi)\} + ... \end{aligned} \tag{14}$$

where

$$E\{\log p(\mathbf{y}|\theta, \sigma^2, \phi, \tau, \psi)\} = -\frac{T}{2}\log(2\pi) - \frac{T}{2}[-\psi(\nu + \frac{T}{2}) + log(\overline{S})] - \frac{\overline{S} - S}{(\frac{\overline{S}}{\nu + \frac{T}{2}})}, \tag{15}$$

$$E\{\log p(\theta)\} = -\frac{k}{2}\log(2\pi) - \frac{1}{2}E[\log|\mathbf{V}|] - \frac{1}{2}[\overline{\theta}^{'}\mathbf{V}^{-1}\overline{\theta} + tr(\mathbf{V}^{-1}\overline{\mathbf{V}})], \tag{16}$$

$$E\{\log p(\sigma^{-2})\} \quad = \quad \nu \log S - \log\Gamma(\nu) - (\nu - 1)[-\psi(\nu + \frac{T}{2}) + log(\overline{S})] - (\frac{S}{\frac{\overline{S}}{\nu + \frac{T}{2}}}), \tag{17}$$

$$E\{\log q(\theta)\} = -(\frac{k}{2} + \frac{k}{2}\log(2\pi) + \frac{1}{2}\log|\overline{\mathbf{V}}_i|), \tag{18}$$

[4]It is seen that the $ELBO$ described below contains many constant terms. These terms need to be dropped from $ELBO$s in VB iterations to speed up the process.

$$E\{\log q(\sigma^{-2})\} = -[\nu + \frac{T}{2} - \log(\overline{S}) + \log(\Gamma(\nu + \frac{T}{2})) - (\nu + \frac{T}{2} - 1)\psi(\nu + \frac{T}{2})]. \tag{19}$$

Notice that

$$\begin{aligned}
&E\{\log p(\mathbf{y}|\theta, \sigma^2, \phi, \tau, \psi)\} + E\{\log p(\theta)\} + E\{\log p(\sigma^2)\} - E\{\log(q(\theta))\} - E\{\log q(\sigma^2)\} \\
&= \frac{k}{2} - \frac{T}{2}\log(2\pi) + \frac{1}{2}\log(|\overline{V}|) - \frac{1}{2}E[\log(|V|)] - \frac{1}{2}[\overline{\theta}' V^{-1}\overline{\theta} + tr(V^{-1}\overline{V})] + \nu\log(\underline{s}) - \log\Gamma(\nu) \\
&- (\nu + \frac{T}{2})\log(\overline{s}) + \log(\nu + \frac{T}{2}) \\
&= \frac{1}{2}\log(|\overline{V}|) - \frac{1}{2}E[\log(|V|)] - \frac{1}{2}[\overline{\theta}' V^{-1}\overline{\theta} + tr(V^{-1}\overline{V})] - (\nu + \frac{T}{2})\log(\overline{s}) + Const.
\end{aligned} \tag{20}$$

Above terms are the baseline *ELBO* discussed in Gefang et al. (2022). The additional terms are as follows:

$$\begin{aligned}
E\{\log p(\tau)\} &= -\log\Gamma(ka) - ka[\int_0^\infty (q(\tau)\log\tau)d\tau] - 0.5\overline{\tau} \\
&= -ka[\int_0^\infty (q(\tau)\log\tau)d\tau] - 0.5\overline{\tau} + Const.
\end{aligned} \tag{21}$$

$$E\{\log p(\psi)\} = \sum_{j=1}^{k}(\log\frac{1}{2} - 0.5\overline{\psi}_j) = -\frac{1}{2}\sum_{j=1}^{k}\overline{\psi}_j + Const. \tag{22}$$

$$\begin{aligned}
E\{\log p(\phi) &= \log\Gamma(ka) - k\log\Gamma(a) - (a-1)\sum_{j=1}^{k}[\int_0^\infty (q(\phi_j)\log\phi_j)d\phi_j] \\
&= -(a-1)\sum_{j=1}^{k}[\int_0^\infty (q(\phi_j)\log\phi_j)d\phi_j]
\end{aligned} \tag{23}$$

$$E\{\log q(\tau)\} = \int_0^\infty q(\tau)\log q(\tau)d\tau. \tag{24}$$

$$E\{\log q(\psi)\} = \sum_{j=1}^{k}\int_0^\infty q(\psi_j)\log q(\psi_j)d\psi_j \tag{25}$$

$$E\{\log q(\phi)\} = \sum_{j=1}^{k} \int_0^{\infty} q(\phi_j) \log q(\phi_j) d\phi_j \tag{26}$$

Terms in $ElBO$ that do not have clear analytical forms can be numerically estimated.

Online Appendix C: Heatmaps of the spatial weights matrices

1 Models where $N = 50$

(a) Sparse W_{n_0} (b) Dense W_{n_0}

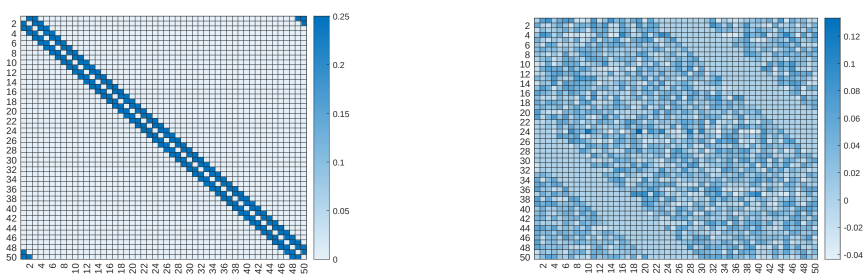
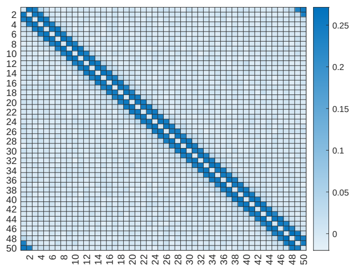
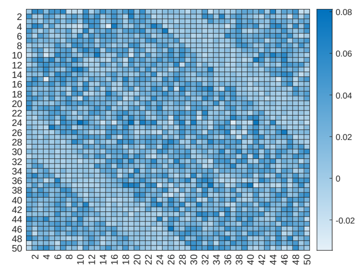


Figure C1: True spatial weights matrices

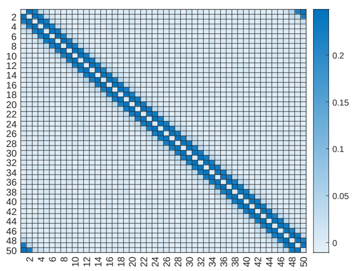
(a) $\lambda_0 = 0.4$, $T = 30$ and sparse W_{n_0}



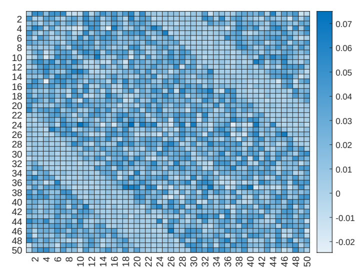
(b) $\lambda_0 = 0.4$, $T = 30$ and dense W_{n_0}



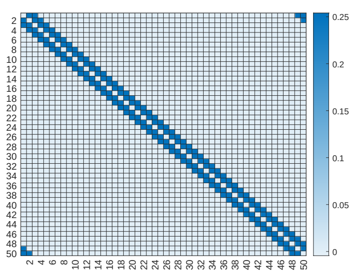
(c) $\lambda_0 = 0.6$, $T = 30$ and sparse W_{n_0}



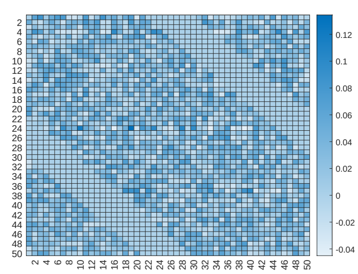
(d) $\lambda_0 = 0.6$, $T = 30$ and dense W_{n_0}



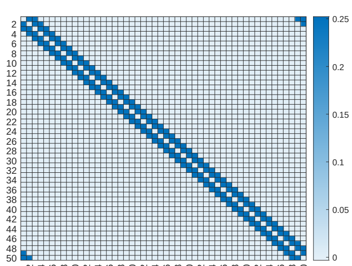
(e) $\lambda_0 = 0.4$, $T = 100$ and sparse W_{n_0}



(f) $\lambda_0 = 0.4$, $T = 100$ and dense W_{n_0}



(g) $\lambda_0 = 0.6$, $T = 100$ and sparse W_{n_0}



(h) $\lambda_0 = 0.6$, $T = 100$ and dense W_{n_0}

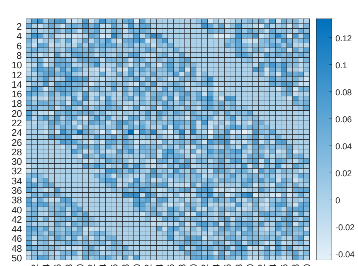
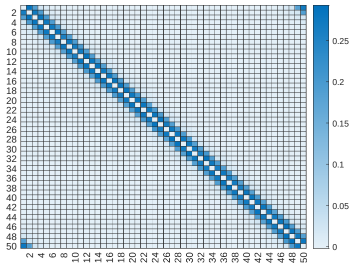
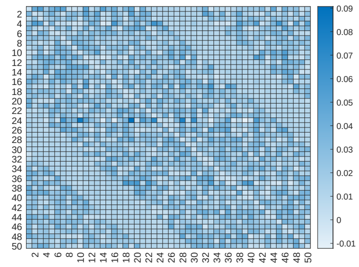


Figure C2: Estimated spatial weights matrices with D-L priors

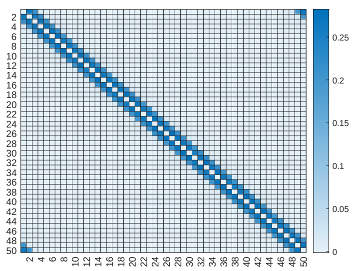
(a) $\lambda_0 = 0.4$, $T = 30$ and sparse W_{n_0}



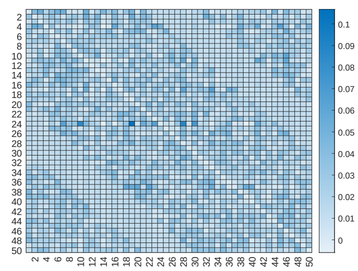
(b) $\lambda_0 = 0.4$, $T = 30$ and dense W_{n_0}



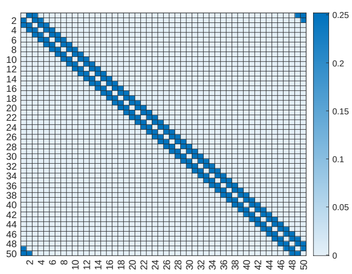
(c) $\lambda = 0.6$, $T = 30$ and sparse W_{n_0}



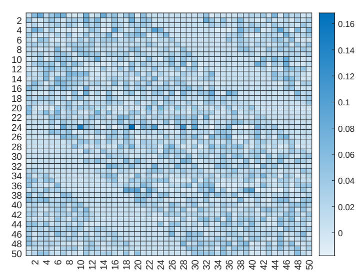
(d) $\lambda = 0.6$, $T = 30$ and dense W_{n_0}



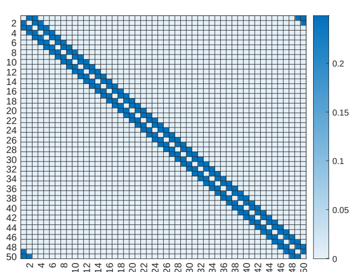
(e) $\lambda_0 = 0.4$, $T = 100$ and sparse W_{n_0}



(f) $\lambda_0 = 0.4$, $T = 100$ and dense W_{n_0}



(g) $\lambda_0 = 0.6$, $T = 100$ and sparse W_{n_0}



(h) $\lambda_0 = 0.6$, $T = 100$ and dense W_{n_0}

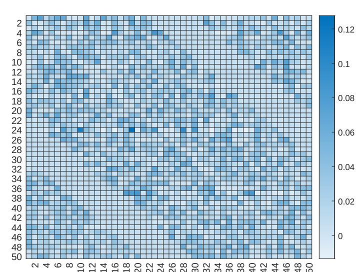


Figure C3: Estimated spatial weights matrices with horseshoe priors

2 Models where $N = 100$

(a) Sparse W_{n_0} (b) Dense W_{n_0}

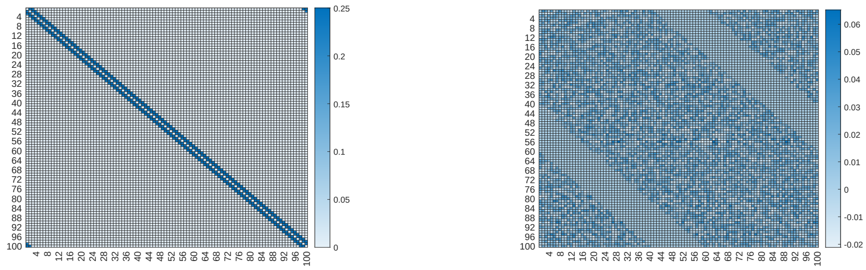
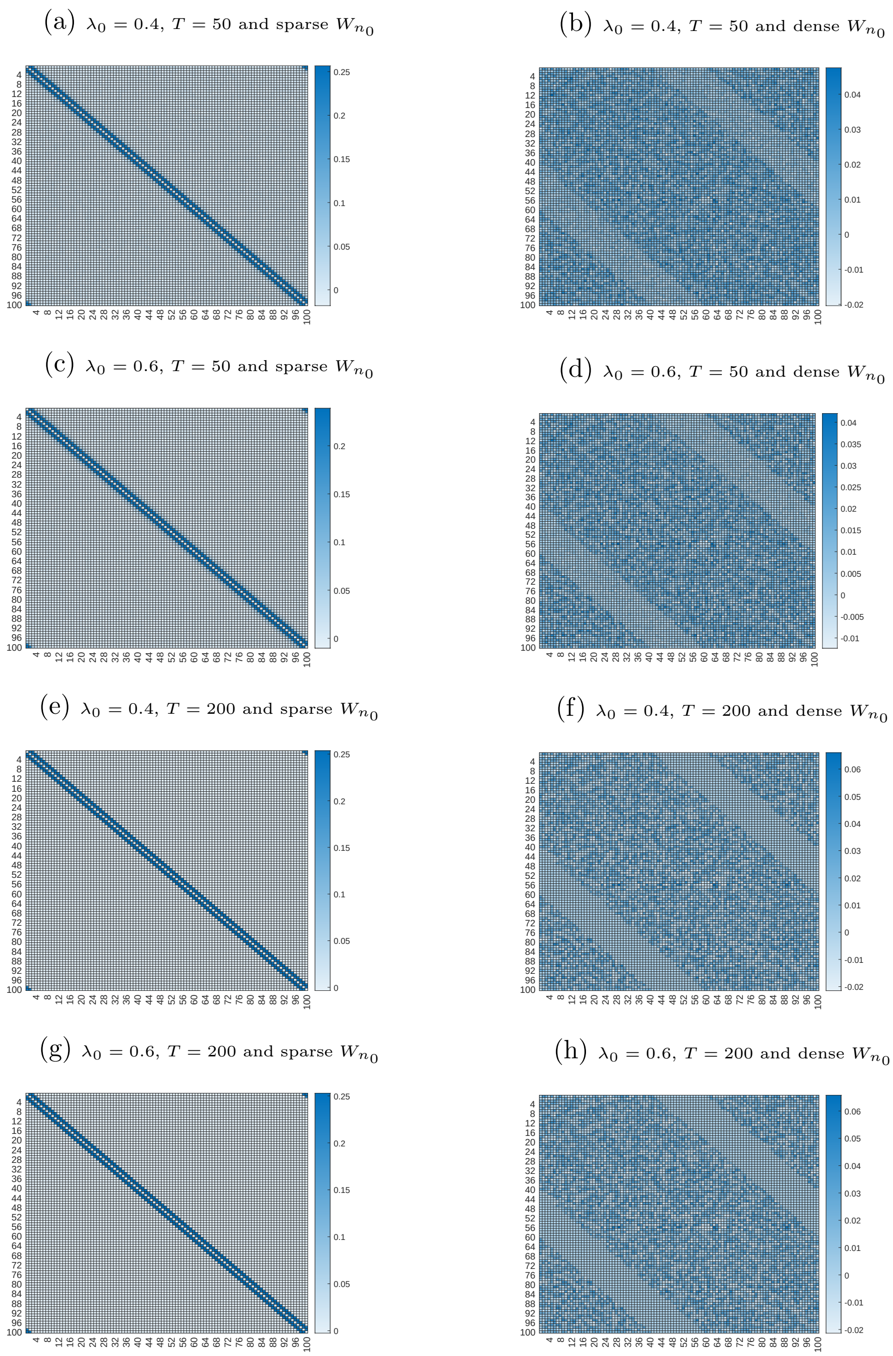


Figure C4: True spatial weights matrices



(a) $\lambda_0 = 0.4$, $T = 50$ and sparse W_{n_0}

(b) $\lambda_0 = 0.4$, $T = 50$ and dense W_{n_0}

(c) $\lambda_0 = 0.6$, $T = 50$ and sparse W_{n_0}

(d) $\lambda_0 = 0.6$, $T = 50$ and dense W_{n_0}

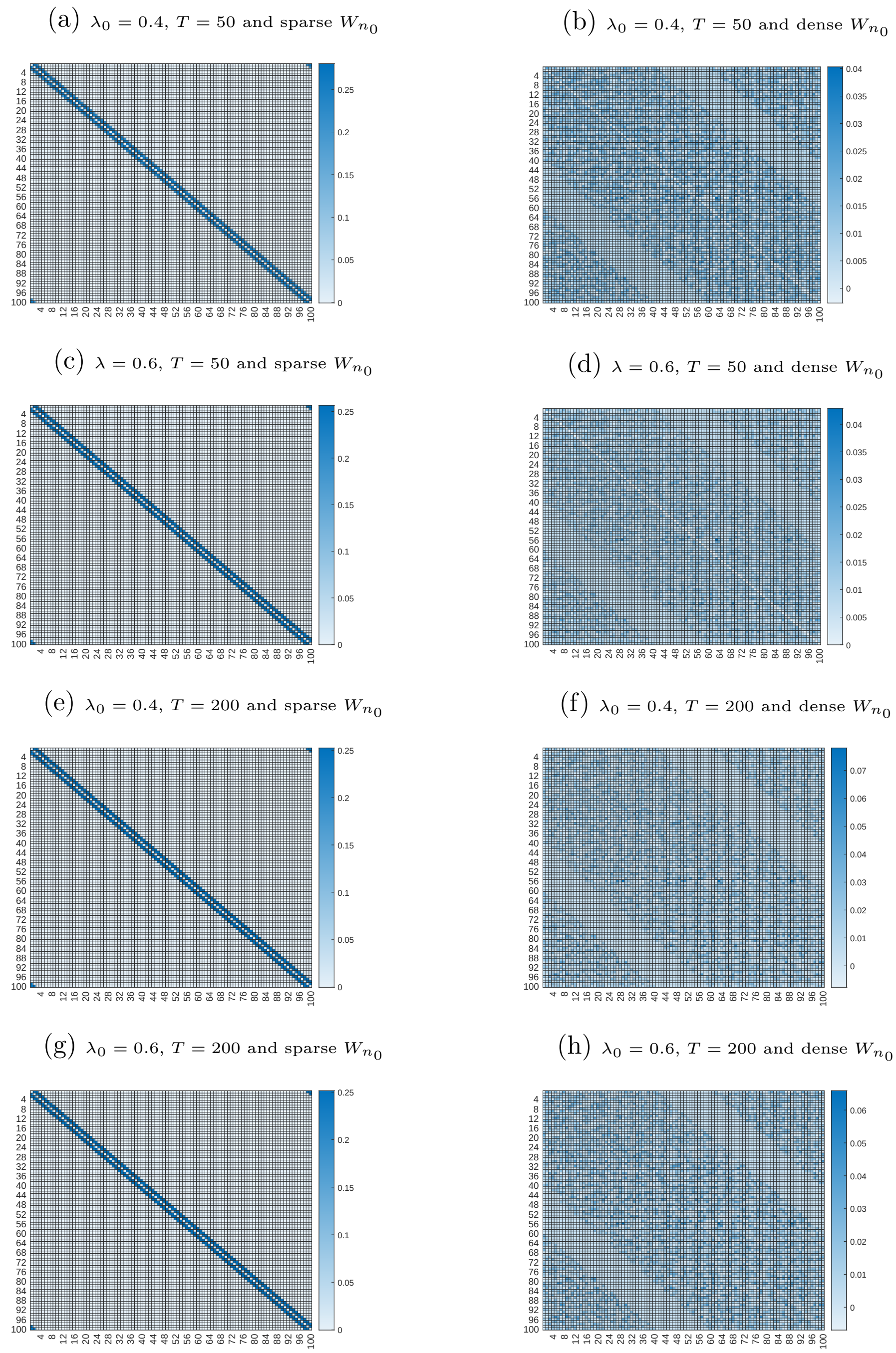
(e) $\lambda_0 = 0.4$, $T = 200$ and sparse W_{n_0}

(f) $\lambda_0 = 0.4$, $T = 200$ and dense W_{n_0}

(g) $\lambda_0 = 0.6$, $T = 200$ and sparse W_{n_0}

(h) $\lambda_0 = 0.6$, $T = 200$ and dense W_{n_0}

Figure C5: Estimated spatial weights matrices with D-L priors



(a) $\lambda_0 = 0.4$, $T = 50$ and sparse W_{n_0}

(b) $\lambda_0 = 0.4$, $T = 50$ and dense W_{n_0}

(c) $\lambda = 0.6$, $T = 50$ and sparse W_{n_0}

(d) $\lambda = 0.6$, $T = 50$ and dense W_{n_0}

(e) $\lambda_0 = 0.4$, $T = 200$ and sparse W_{n_0}

(f) $\lambda_0 = 0.4$, $T = 200$ and dense W_{n_0}

(g) $\lambda_0 = 0.6$, $T = 200$ and sparse W_{n_0}

(h) $\lambda_0 = 0.6$, $T = 200$ and dense W_{n_0}

Figure C6: Estimated spatial weights matrices with horseshoe priors

Online Appendix D: More plots on VB estimates vs. MCMC posterior draws

In this appendix, we present the MCMC posterior draws, VB estimates, and true parameters corresponding to the non-zero elements of the spatial weight matrices. We use W_1, $n=?$ to indicate the non-zero element preceding the diagonal entry in the n^{th} row of the spatial weight matrix, and W_2, $n=?$ to indicate the non-zero element that following the diagonal entry in the n^{th} row of the spatial weight matrix. Each subfigure displays the histogram of MCMC posterior draws (in light blue), the VB point estimate (in red), and the true value of the parameter (in dark blue).

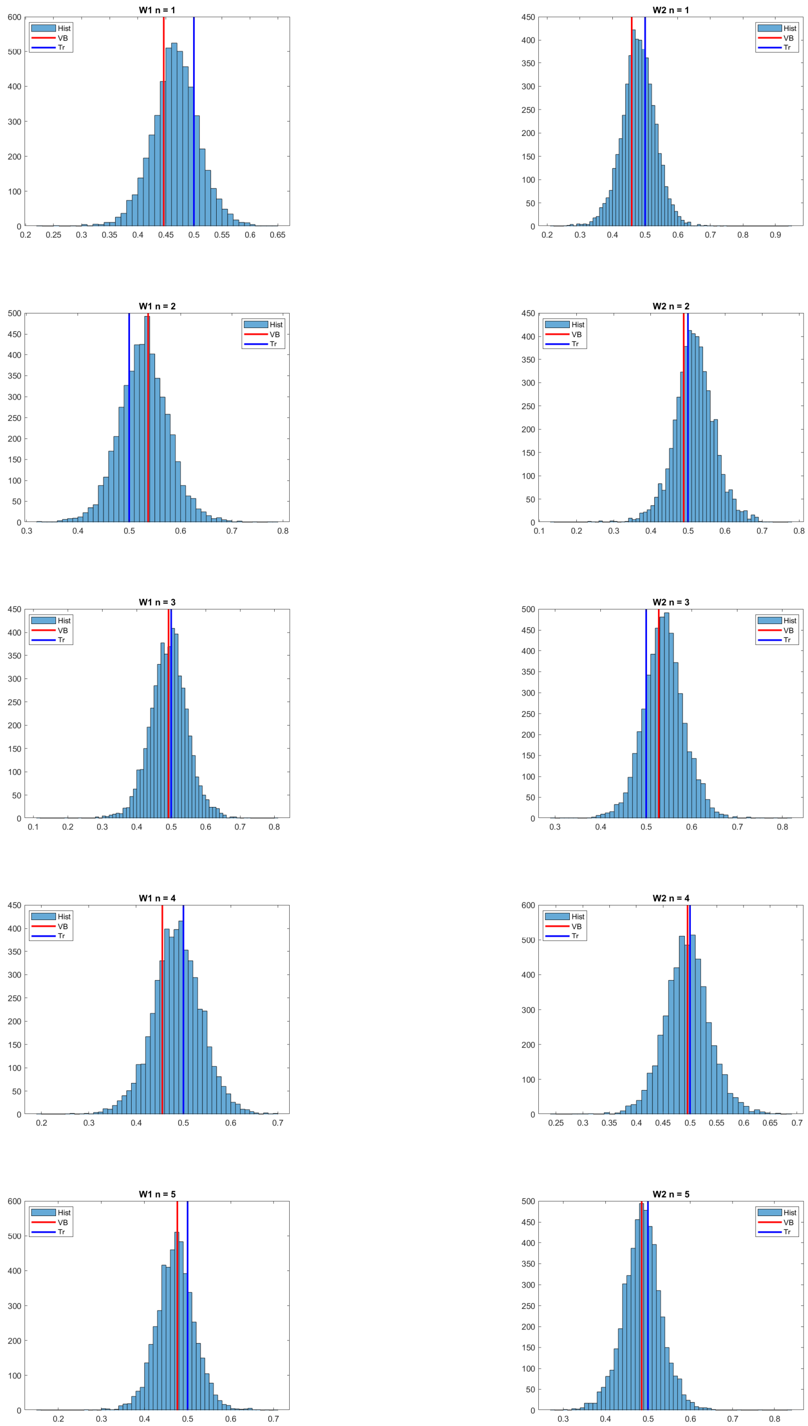


Figure D1: Plots of MCMC posterior draws against VB estimates and true parameter values, with D-L priors

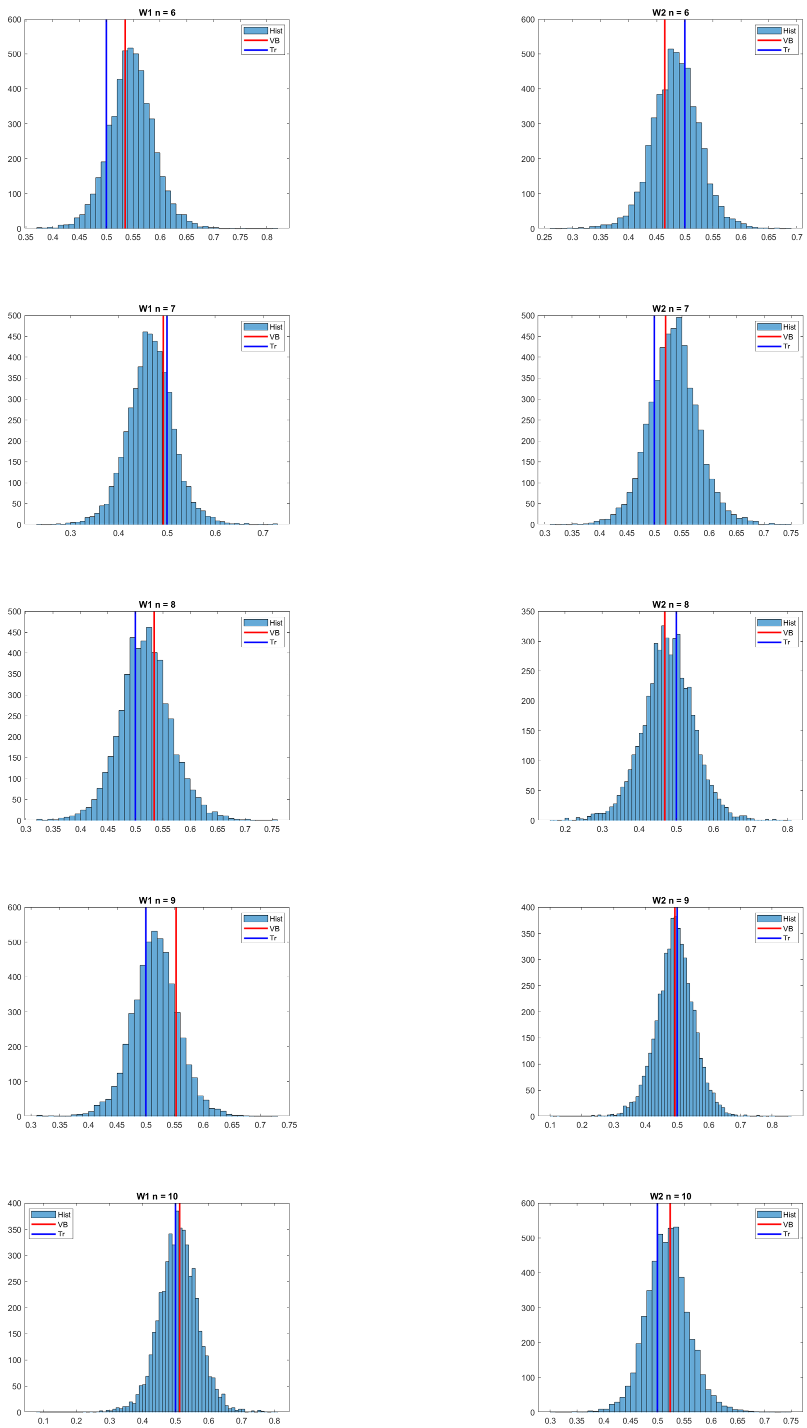


Figure D2: Plots of MCMC posterior draws against VB estimates and true parameter values, with D-L priors

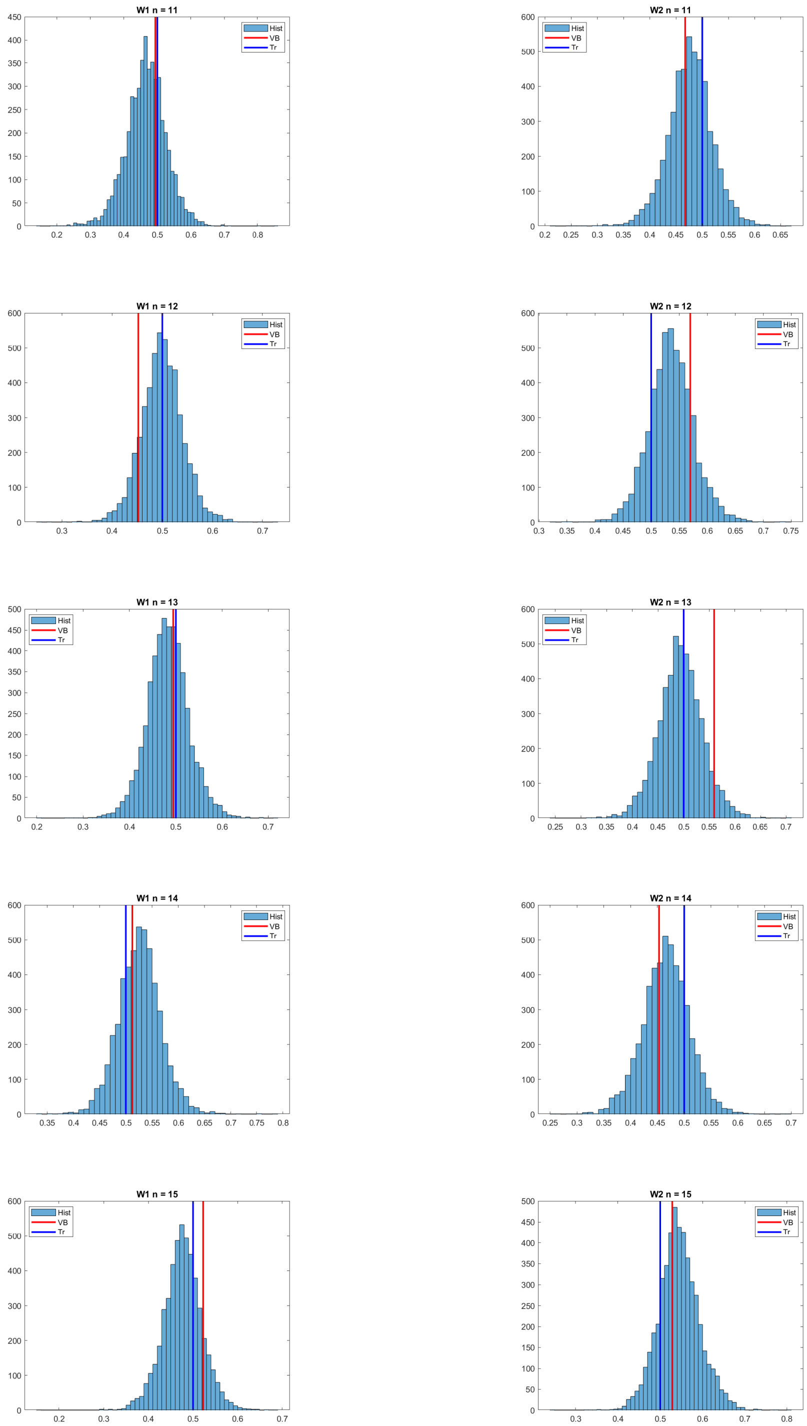


Figure D3: Plots of MCMC posterior draws against VB estimates and true parameter values, with D-L priors

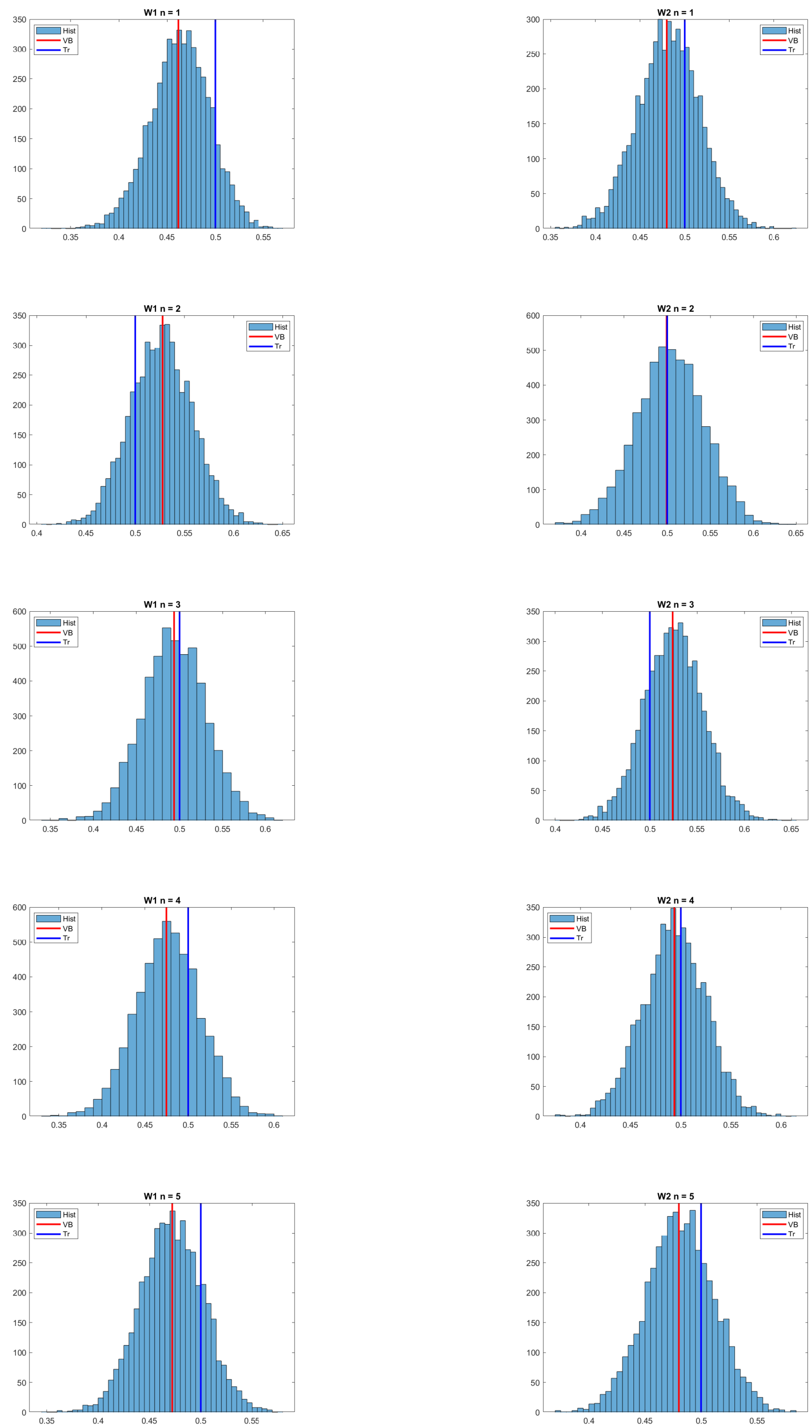


Figure D4: Plots of MCMC posterior draws against VB estimates and true parameter values, with horseshoe priors

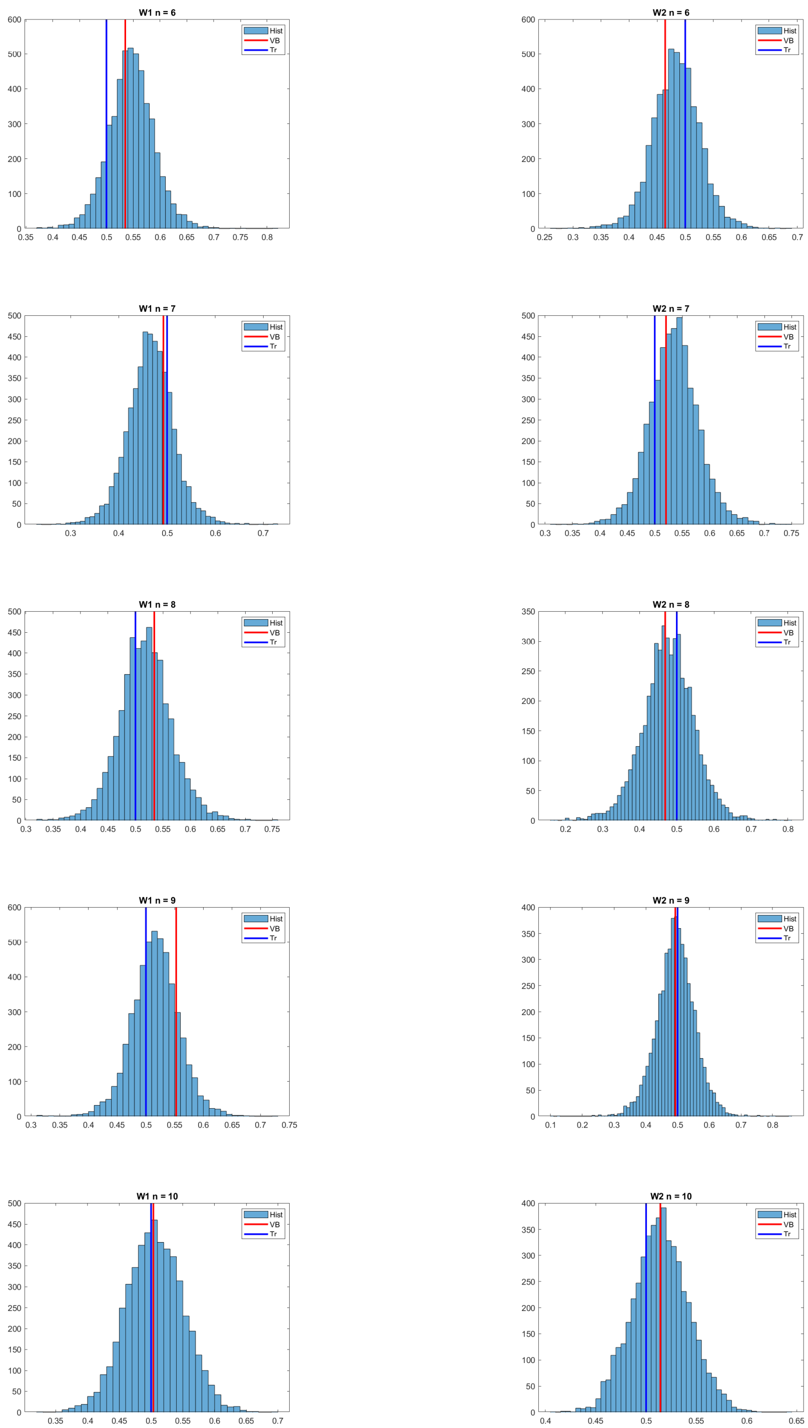


Figure D5: Plots of MCMC posterior draws against VB estimates and true parameter values, with horseshoe priors

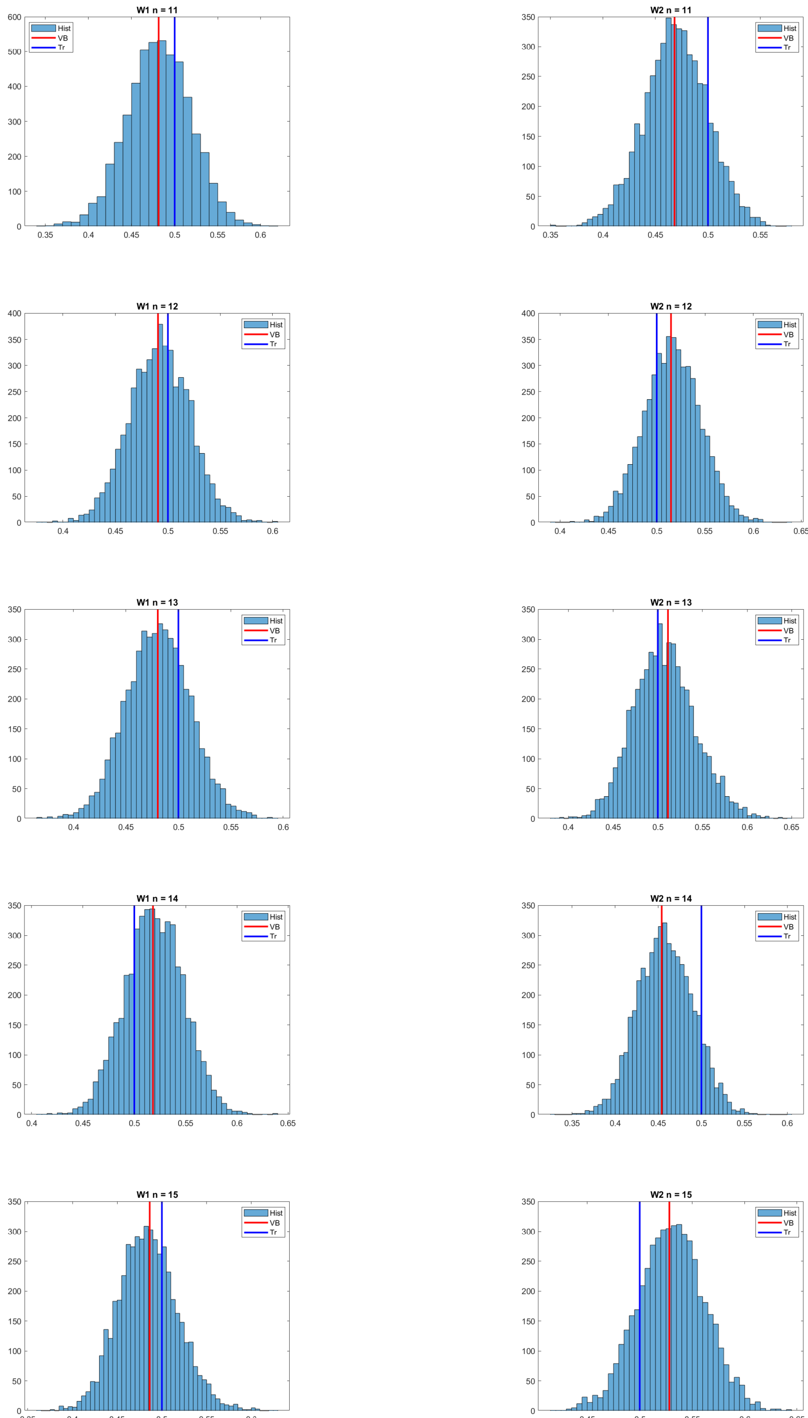


Figure D6: Plots of MCMC posterior draws against VB estimates and true parameter values, with horseshoe priors

Online Appendix E: Data Sources

(a) Spreads: monthly data from Statistical Data Warehouse, European Central Bank

(b) Ratings: monthly data from Fitch, Standard and Poor's and Moody's

(c) The ratio of government debt to GDP: quarterly data from Thomson Reuters Datastream

(d) Real GDP growth: quarterly data from Thomson Reuters Datastream

(e) Current account balance as a percentage of GDP: Either monthly or quarterly data from Thomson Reuters Datastream

(f) Harmonised Consumer Price Index: monthly data from Thomson Reuters Datstream

(g) Political uncertainty: quarterly data from IFO World Economic Survey. Two variables have been used in order to create the variable political uncertainty. 1) The present climate for foreign investors, political stability (which was discontinued), which was updated with information from the new variable, 2) Political instability, using an appropriate mapping for the reported values of each variable.